

# 디지털자료실지원센터 종합목록 데이터 품질평가 및 관리 방안

## Evaluation and Quality Control of Data in the Digital Library System

최 인 숙(In-Sook Choe)\*

### 목 차

- |                             |                     |
|-----------------------------|---------------------|
| 1. 서 론                      | 4. 종합목록데이터베이스의 품질평가 |
| 1. 1 연구의 목적                 | 4. 1 적중률            |
| 1. 2 연구의 방법 및 제한점           | 4. 2 중복성            |
| 2. 선행연구                     | 4. 3 완전성            |
| 3. 서지레코드의 구조 및 데이터 입력<br>규칙 | 4. 4 정확성            |
|                             | 5. 결 론              |

### 초 록

이 연구의 목적은 디지털자료실지원센터 종합목록데이터베이스의 품질을 평가하고 관리하기 위한 방안을 제시하는 것으로 적중률, 중복성, 완전성, 정확성을 기준으로 삼아 데이터의 품질을 분석하였다. 데이터베이스의 적중률은 매우 높은 편이었으나 동일 저작에 대해 상이한 레코드가 다수 존재하는 중복성은 심각한 수준이었다. 또한 입력해야할 항목에 대한 인식이 매우 상식적인 수준에 그쳐 레코드의 완전성 정도는 48.12%이었으며, 정확성 분석 결과 92%에 해당하는 레코드에서 다양한 오류가 발견되었다. 품질관리를 위한 제언과 더불어 오류 유형 분석을 통해 오류의 원인과 방지대책을 살펴보았다.

### ABSTRACT

This study intends to evaluate the quality of the Digital Library System DB and to suggest methods for its quality control. The evaluation criteria are hit rates, redundancy, completeness and accuracy. In spite of high hit rates excessive records representing one work resulted serious redundancy. The average completeness rate of records was 48.12% due to low level of description. The Analysis of accuracy showed various errors in most of records corresponding to 92%. Emphasis on analysing the errors in detail detected the causing factors and suggested practical guidelines for school libraries' catalogers.

키워드: 디지털자료실지원센터, DLS, 품질평가, 품질관리

\* 연세대학교 시간강사(ischoe@hanafos.com)  
논문접수일자 2004년 8월 10일  
게재확정일자 2004년 9월 10일

## 1. 서론

디지털자료실지원센터(Digital Library System, 이하 DLS라 함)는 시·도교육청에 호스팅 서버를 설치하여 관내 학교도서관의 도서관리업무를 자동화하고 독서교육 지원 기능을 통합적으로 서비스하는 웹 기반 학교도서관 정보시스템이다. DLS는 시·도교육청의 전산실 서버에 학교도서관 업무지원 시스템, 공동목록 시스템, 독서 및 정보 활용교육 시스템이 탑재되어 있기 때문에 단위 학교에서는 별도의 자료관리 프로그램을 구매하지 않더라도 인터넷에 연결된 PC만으로 다수의 학교도서관이 동시에 사용할 수 있다. 개별 학교도서관의 도서 구입이나 목록 정리, 자료 검색, 학생들의 대출, 반납, 통계 처리 업무는 교육청 시스템을 통해 통합적으로 이루어지게 된다. 특히 목록 정리에 있어 교육청에 구축된 종합 목록을 활용하면 개별 학교마다 소장 자료를 모두 입력해야 했던 작업의 중복 문제를 획기적으로 해결할 수 있다.

2001년 5월 교육인적자원부는 '디지털자료실지원센터 구축 운영' 사업에 대한 기본 계획을 수립하여 부산광역시 교육청에 학교도서관지원시스템으로서 DLS-I을 시범 구축하였고 2002년에는 충청북도 교육청에 DLS-II를 시범적으로 구축·운영한 이래 현재 전국 16개 교육청에 시스템이 구축되어 활용을 추진하고 있다. DLS는 2001년 데이터베이스 구축을 시작한 이래 3년 여 짧은 기간 동안에 무려 6,477,086 건의 서지레코드를 포함하는 대규모 데이터베이스로 성장하였다(교육인적자원부, 한국교육학술정보원 2003). 그러나 양적

성장에도 불구하고 시스템의 불안정과 데이터의 조악한 품질로 인해 그 유용성은 이용자들 사이에서 큰 호응을 얻지 못하고 있는 실정이다(학교도서관을 살리는 교사들 2004).

### 1. 1 연구의 목적

본 연구의 목적은 DLS의 기본 방향을 제시하는 원형이자 가장 많은 양의 데이터를 보유한 부산광역시 교육청 DLS(<http://dls.busanedu.net/>)의 데이터베이스를 대상으로 데이터의 품질을 분석하고 그 결과를 기본 자료로 삼아 전국 16개 DLS의 데이터 품질을 개선할 수 있는 방안을 제시하는 것이다.

부산광역시 교육청 DLS는 2003년 9월 현재 목록 데이터베이스에 1,243,265 건이 구축되어 있다(교육인적자원부, 한국교육학술정보원 2003). 그러나 16개 DLS 중에서 가장 많은 데이터를 보유하고 선도적 역할을 하는 시스템임에도 불구하고 데이터베이스의 품질이나 시스템의 안정성 면에서 아직 신뢰감을 주지는 못하고 있다. 각 학교도서관이 기 구축한 데이터베이스를 단기간에 통합하는 과정에서 생겨난 중복 및 오류레코드로 인한 불가피한 일이라고 이해할 수는 있으나 서지데이터의 품질검증 및 관리는 하루 빨리 해결해야만 하는 시급한 사안이며 보다 적극적인 방안이 절실한 시점이다. 특히 목록 규칙을 잘 모르는 비전문가가 학교도서관을 관리하는 경우가 많고 입력자가 자주 교체되어 일관성 있는 작업을 하기 어렵다는 실정을 감안할 때 실무자들이 유의할 사항에 대해 살펴보는 일은 매우 중요하다고 하겠다. 현재 부산광역시 교육청

DLS시스템의 운영실태에 대한 연구(이병기 2003)와 정보시스템 소개 및 기반 기술에 대한 연구(함명식 2002)가 있지만 데이터베이스의 품질평가는 본격적으로 행해지지 않은 상태이다. 따라서 본 연구에서는 부산광역시 교육청 DLS 종합목록 데이터의 품질을 평가하여 품질 관리를 위한 제언을 하고 각 필드의 오류 유형 분석을 통해 오류의 원인을 살펴본 후 오류 방지를 위해 입력 실무자들이 유의할 사항을 제시하고자 한다.

## 1. 2 연구의 방법 및 제한점

데이터베이스를 평가하는 데 있어 무엇보다 중요한 문제는 평가 기준의 설정이라 할 수 있다. 본 연구에서는 데이터의 품질을 중심으로 평가하기 위해 적중률, 중복성, 완전성, 정확성을 기준으로 삼았다. 적중률은 데이터베이스에 수록되는 자료의 규모를 나타내기 위한 기준으로 일정 수의 자료를 검색했을 때의 검색 성공률로 측정하였으며, 중복성은 동일 자료를 표현하는 복수 레코드의 비율을 나타내기 위하여 동일한 저작에 대해 검색되는 레코드의 건수로 측정하였다. 완전성은 레코드에 포함된 데이터의 내용이 원자료의 중요한 속성을 표현하고 있는 상세성의 정도를 표현하기 위한 기준으로 각 서지에 대해 기술해야할 항목의 총수 가운데 실제로 기술한 항목수의 비율을 측정하였고, 정확성은 데이터가 실제 값과 동일한 정도를 나타내기 위해 각 항목별로 기술 내용의 철자나 데이터값 오류, 데이터값 누락, 혹은 목록규칙 위반에 해당하는 레코드의 발생율을 계산하여 측정하였다. 본 연구에서는

정확성 평가에 많은 비중을 두었는데 오류를 입력 요소별로 분류하고 분석함으로써 원인을 좀 더 자세히 살펴보고 학교도서관의 입력 담당자들이 오류 방지를 위해 어떤 점에 유의해야 하는지를 파악하기 위함이다.

현재 DLS에 수록된 데이터가 대부분 단행본임을 고려하여 연구 대상 자료 유형은 단행본으로 제한하였고, 2004년 6월 1일에서 6월 10일까지 10일 동안 본 연구자가 부산광역시 교육청 DLS에 일반 이용자 자격으로 접속하여 검색한 후 결과물을 출력·분석하였다. 검색 대상 자료는 부산광역시 교육청에서 제시한 '2004 독서인증시스템과 연계한 권장도서 목록'에 수록된 중학생용 단행본 100종, 학교 현장에서 독서교육을 실천하는 교사들의 모임인 '책으로 따뜻한 세상을 만드는 교사들(이하 책따세라 함)'에서 제시한 '2003년 겨울, 책/따/세가 청소년에게 권하는 책 목록(분야별)'에 수록된 단행본 33종, 책따세가 중학생에게 권한 책에서 쉬운 책을 따로 뽑아 방학용으로 제시한 '2000년 여름부터 2003년 겨울까지, 책따세에서 함께 읽고 함께 뽑은 책'에 수록된 단행본 80종, 서울 N중학교에서 제시한 '필독도서 및 권장도서 목록'에 수록된 단행본 66종으로 총 279종이다.

조사 대상 자료가 모집단을 대표하기에 충분한 규모가 아니므로 품질관리를 위한 제언과 오류방지를 위해 입력실무자들이 유의할 사항을 개략적으로 파악할 수 있었을 뿐 전체적이고 완전한 시각은 다소 결여되어 있음을 제한점이라 할 수 있다.

## 2. 선행연구

국내외에서 데이터베이스의 평가에 관심을 갖는 많은 연구자들이 품질평가 기준을 제시한 바 있다. 보편적인 기준으로 활용되고 있는 SCOUG(The Southern California Online User's Group)의 평가법은 사용자의 관점을 품질평가 기준에 적극 반영한 것으로서 서지 데이터베이스나 전문데이터베이스를 질적 측면에서 평가할 수 있는 11가지 측정기준 즉 일관성, 범위, 시기적절성, 비용의 가치, 접근성, 시스템의 형태, 사용의 용이성, 다른 데이터베이스와의 통합성, 검색 결과물의 제공형태, 다큐멘테이션, 이용자의 호응도 등을 제시하였다(Basch 1990). Jasco는 CD-ROM 데이터베이스와 온라인 데이터베이스를 대상으로 내용의 질적 평가 요소 및 네트워크적인 평가 요소로 권위, 최신성, 주제범위, 내용, 접근성, 이용자 지원, 비용, 호스트 성능, 검색 소프트웨어, 통신망 등 10가지를 제시하였고(Jasco 1995), AT & T Bell 연구에서는 데이터베이스 품질을 위해 정확성, 최신성, 완전성, 통합성 등 4가지 기준만 간단히 제시하기도 하였다(Fox 1994).

국내의 선행연구로 한국데이터베이스진흥센터에서는 데이터 품질과 서비스 품질로 나누어서 접근하였다. 데이터 품질은 데이터 자체의 바람직한 정도와 데이터 양, 전문성을 의미하는 것으로 정확성, 완전성, 최신성, 수록범위, 전문성을 기준으로 제시하였고 서비스 품질은 사용자에게 서비스되는 과정에서의 매체, 수단, 방법, 공급처 지원 등의 바람직한 상태를 의미하는 것으로 검색성, 접근성, 사용

용이성, 사용자 지원성, 비용, 네트워크 및 하드웨어 등을 기준으로 제시하였다(한국데이터베이스진흥센터 2000).

한국과학기술정보연구원(2001), 한국교육학술정보원(2002), 김석영(2002), 김성희(2003), 이만수 등(2003), 김상준(2003) 등 많은 기관과 연구자들이 국내외 데이터베이스를 대상으로 다양한 기준과 측정방법을 제시한 바 있지만 본 연구에서는 분석 대상으로 삼고 있는 데이터베이스가 분담 편목 방식으로 분산 구축한 데이터베이스이기 때문에 유사한 경우를 집중적으로 살펴보았다. 이제환은 분산체제로 구축된 해외 데이터베이스인 OCLC Union Catalog의 품질 관리 전략을 참고하면서 분산체제로 구축된 국내의 SATURN 데이터베이스의 품질을 검증하고 관리방안을 제시하였다. 검색실험을 통해 표본 레코드를 수집한 후 국내 문헌에 대한 서지레코드와 해외문헌에 대한 서지레코드의 비율, 자체 제작 레코드의 비율, 레코드의 최신성 및 갱신주기, 레코드의 중복률, 레코드 구조의 일관성, 데이터필드의 적합성, 수록데이터의 완전성 등 7가지 항목을 중심으로 품질검증을 하였고 이용자의 관점에서 본 유용성을 측정하기 위해 인용문헌분석법과 인터뷰를 시행하기도 하였다(이제환 1998). 또 한국교육학술정보원(이하 KERIS라 함)의 종합목록데이터베이스에 대해 사용자인 현직사서들과 최종이용자 그룹이 느끼는 유용성을 판단하기 위해 포괄성, 배타성, 최신성, 중복성, 일관성, 완전성을 기준으로 채택한 후 이 6개 기준을 실질적으로 측정하기 위한 지표를 개발하여 데이터베이스의 품질을 검증하였다(이제환 2002). 이응봉

등은 데이터베이스에 수록된 데이터의 품질과 데이터베이스 제공 서비스의 품질 측정을 위해 총 18개의 항목을 사용하여 과학기술분야 데이터베이스의 품질을 평가하고 개선방안을 제안하였다(이응봉 외 2001). 이 세 연구에서는 중앙기관의 철저한 감독과 관리가 필요하며 수록 데이터에 대한 검증과 오류데이터의 수정을 정기적으로 시행할 수 있는 장치가 필요하다는 제언을 하였는데 분산구축방식의 공동목록 데이터베이스를 대상으로 하는 본 연구에 대해 시사하는 바가 많았다.

종합목록 데이터의 품질을 평가하고자 하는 본 연구에서 참고할 또 다른 유형의 선행연구는 레코드의 기술 부분에 주목하여 데이터의 오류 유형을 분석하고 편목의 중요성에 대한 인식을 제고하는 연구들이다. 조순영은 KERIS 종합목록의 학위논문 서지데이터를 대상으로 각 회원 기관이 통일된 입력기준 없이 구축된 데이터베이스를 통합하는 과정에서 오류가 발생할 수밖에 없음을 지적하면서 오류데이터의 유형을 분석함으로써 데이터의 품질을 측정하고 기계적으로 오류데이터를 색출할 수 있는 방안을 제시하였다. 오류데이터가 발생하는 주요 원인으로는 서지레코드의 입력 오류, MARC의 사용 오류, 목록규칙의 적용 오류 등을 지적하였다(조순영 2002). 윤정옥 또한 기존의 데이터베이스 평가가 구조나 검색의 효율, 인터페이스에 치중하고 콘텐츠인 데이터의 품질 평가나 개선 문제에 주의를 기울이지 못했다는 인식 하에 서지레코드의 기술 부분을 집중 분석하였다. 한국과학기술정보연구원(KISTI)의 과학기술 연속간행물 종합목록데이터베이스를 대상으로

AACR2와 KORMARC 포맷 등의 편목규칙과 기술표준을 기준으로 하여 각 필드별로 발생하는 오류의 유형을 분석하였는데 편목규칙과 기술표준을 따름으로써 오류를 수정하고 데이터베이스의 품질향상에 기여할 수 있다고 주장하였다(윤정옥 2003). 목록에 수록된 내용이야말로 이용자들이 가치를 두고 있는 부분임을 강조하면서 지금까지 목록이 추구해왔던 목록생산자와 목록전문가의 관점을 지양하고 이용자의 입장에서 보다 실질적인 유용성을 증대해 나갈 수 있는 방향으로의 패러다임 전환을 강조한 노지현의 연구도 눈길을 끈다. 내용적 유용성을 측정하기 위한 기준으로 포괄성(관련 자료를 포괄적으로 탐색해 낼 수 있는가), 상세성(적합한 자료를 식별하는 데 도움이 되는 데이터를 상세하게 수록하고 있는가), 완전성(원자료의 속성을 완전하게 기술하고 있는가), 정확성(원자료의 물리적인 속성을 정확하게 반영하고 있는가) 등을 적용하여 6개 대학교도서관의 목록을 대상으로 검증한 결과 네 부분 모두에서 심각한 문제점을 발견하였으며 특히 수록된 내용의 상세성이 가장 취약하다고 지적하였다(노지현 2003).

### 3. 서지레코드의 구조 및 데이터 입력 규칙

각 학교도서관에서는 종합목록에서 조회된 서지데이터 혹은 신규 생성된 서지를 소장목록에 추가하거나 수정, 삭제를 함으로써 각 학교의 데이터를 관리한다. 각 학교에 자료가 입수되면 우선 자관검색을 통해 소장 도서인지

복본조회를 하고 소장 자료에 없다면 종합목록에 구축된 자료가 있는지 조회한다. 종합목록에 자료가 있다면 자관으로 서지반입 후 등록번호, 청구기호, 소장위치 등을 입력하여 각 학교의 도서로 추가하고 종합목록에 자료가 없다면 새로운 서지로 신규등록하고 자관 소장정보를 등록한다.

DLS 센터에서는 단위학교에서 등록하는 데이터에 대해 중복검사를 하고 각 시·도교육청에 위치한 종합목록데이터베이스에 저장하여 종합목록데이터를 공동 활용할 수 있도록 한다. 중복검사 과정에서 ISBN, 서명, 저자, 출판사, 출판년도, 판사항, 페이지수 등을 비교하여 동일한 서지임이 확인되면 서지병합처리를 하여야 한다. 서지병합 처리를 할 때는 주기사항이나 표지이미지 등이 상세하게 기술되어 서지 데이터의 질이 높은 쪽에 병합되도록 하여야 한다. 또 저자 전거데이터를 관리하여 동명이인의 저자명을 전거 통제하고 중복입력을 방지하도록 한다.

자료를 신규 등록할 때는 기본서지를 먼저 입력하고 상세사항은 '소장/확장서지'에서 각 항목을 눌러 추가 입력하게 되어있다. 기본사항과 상세사항으로 구분하여 레코드의 구조를 나타내고 입력시 지켜야할 서지입력규칙을 기본서지에 한하여 간단히 설명하면 <표 1>과 같다.

소장정보등록은 소장/확장서지정보에서 등록번호와 청구기호, 구입가, 인쇄년도, 입수유형, 배가위치, 도서상태, 딸림정보를 입력하면 된다. 구입가는 실제로 도서를 구입한 가격을 입력하고 인쇄년도는 출판년이 아닌 가장 최근에 인쇄된 년도를 입력한다. 입수유형은 도

서가 입수된 유형으로 구입, 기증, 교환 중에서 선택한다. 배가위치는 도서가 입력되고 소장되어질 장소를 선택한다. 도서상태는 현재 도서의 상태를, 딸림정보는 출판물에 부속되어 출판된 자료로서 지도, 악보, 카세트 테이프 또는 해설서 등을 입력하면 된다.

## 4. 종합목록데이터베이스의 품질 평가

### 4.1 적중률

종합목록데이터베이스의 적중률을 파악하기 위해 부산광역시 교육청 DLS에서 단행본 279종을 대상으로 서명, 저자명, 출판사를 입력한 후 검색된 결과를 적중률 측정에 이용하였다. 검색결과 부산광역시 교육청 제시 자료의 검색 성공율은 98%이고 책따세 2003년 겨울 자료와 책따세 방학 자료, 서울 N중학교 자료의 검색 성공율은 모두 100%로 측정되었다. 검색대상 자료 279종에는 한국문학, 외국문학, 인물관련 자료, 교과관련 자료가 두루 포함되어 있으며 주제 분야도 치우침 없이 망라되어 있다. 따라서 종합목록데이터베이스에 수록된 자료는 매우 높은 적중률을 보이며 중학생을 위한 추천 자료의 범위를 충분히 만족시키는 수준으로 판단된다.

### 4.2 중복성

종합목록데이터베이스의 중복성을 파악하기 위해 부산광역시 교육청에서 제시한 목록

〈표 1〉 서지레코드의 구조 및 입력규칙

기본사항	상세사항	서지입력규칙
ISBN		도서마다 부여되는 일련번호 10자리를 '-' 없이 입력한다.
본서명		권차를 뺀 순수한 도서명만 입력하며 서명 등록 시 실물자료의 표제지, 표지, 판권지순으로 참조한다.
권차		1부 2책 이상이거나 다권본의 도서에서 그 순서를 표시한 숫자나 단어를 입력한다.
권차서명		다권본의 경우 본서명 외에 권차에 딸린 서명을 입력한다.
	총서명	
	부서명	
저자		책의 저자나 편집자를 표지 또는 판권지에 기입된 그대로 입력한다. 동양저자의 경우 띄어쓰기 없이 기입하고, 서양저자의 경우 이름과 성 사이를 띄어쓰기한다.
	생몰년	
	저자전거	
	역할어	선택항목에서 선택하며 선택항목에 없는 역할어는 직접 기입한다.
출판사		도서가 출판된 회사를 입력한다.
	발행지	
발행년		해당관의 인쇄연도가 아닌 출판년도를 입력한다.
페이지수		자료의 전체 페이지수를 입력한다.
	삼도	
크기		도서의 세로길이를 올림해서 cm 단위로 입력한다.
판사항		증보판, 개정판, 제3판 등 자료의 판차사항을 입력한다. 초판은 입력하지 않는다.
가격		도서의 정가격을 입력한다.
	자료형태	
분문기술언어		도서에 쓰여진 언어형식을 선택한다.
	저작권	
이용대상		도서를 이용하는 적정대상을 선택한다.
	분류명키워드	
분류번호		해당주제에 해당하는 분류번호로서 KDC에 준해 기입한다.
간략내용		도서의 내용이나 초록을 간략하게 입력한다.
표지이미지		
	초록	
	주기	
	연관저록	
	URL	
	목차	

을 대상으로 동일한 저작에 대해 검색되는 레코드의 건수를 집계하였다. 이 목록에 제시된 자료는 100종이지만 권차와 권차서명이 있는 자료를 별개의 것으로 간주하여 총 129권으로 처리하여 검색하였고 그 결과 963건이 검색되었다. 이는 하나의 자료에 대해 7.46개의 상이한 레코드가 존재한다는 의미로서 이들이 중복레코드임에도 불구하고 별개의 레코드로 처리된 이유는 각 학교도서관에서 입력한 레코드의 구조에 일관성이 없고 입력된 데이터가 정확하지 않기 때문이라 할 수 있다. 상세한 내용은 4.3과 4.4에서 분석할 예정이다. 중복성이 7.46이라는 결과는 상당히 심각한 문제점이며 DLS시스템의 중복체크 기능에 대해 재고해 볼 여지가 있음을 보여준다고 하겠다.

### 4. 3 완전성

종합목록데이터베이스의 완전성을 파악하기 위한 자료는 부산광역시 교육청 제시 '2004 독서인증시스템과 연계한 권장도서목록' 중 한국문학, 외국문학, 교과관련 도서가 고루 포함되면서 일서, 원서명 있는 서양서, 권차와 권차서명이 있는 도서, 관청이나 잡제 등 기타 서명이 있는 도서, 개정판, 총서, 공저자 저작 등 목록 기술상의 기본적인 테크닉을 볼 수 있는 도서로 하였다. 가능한 검색건수가 10건 이상인 자료를 선별한 결과 총 13종 177레코드가 선정되었고 상세목록을 분석하였다. 선정된 자료는 <표 2>에 제시하였다.

레코드의 완전성을 측정하기 전에 우선 항목별 입력률을 살펴보았다. 항목별 입력률은 각 항목을 포함해야만 하는 레코드의 총수 중 해당 항목이 실제로 입력된 레코드수의 비율

<표 2> 완전성 측정을 위해 선정된 자료

	서명	저자	발행처	검색건수
한국문학 4종	7년간의 실수	김주환	나라말	16
	흥어	김주영	문이당	8
	모두 아름다운 아이들	최시한	문학과지성사	11
	그 많던 싱아는 누가 다 먹었을까	박완서	웅진닷컴	24
서양문학 4종	동물농장	조지 오웰	청목	4
	나무	베르나르 베르베르	열린책들	10
	얼굴 빨개지는 아이	장 자끄 상빠	열린책들	12
	난 죽지 않을테야	세르쥬 베레르	문원	12
일본문학 2종	나 이제 외톨이와 안녕할지 몰라요	하이타니겐지로	사계절	17
	창가의 토토	구로야나기테츠코	프로메테우스	13
교과관련 3종	대한민국사 1-2	한홍구	한겨레신문사	17
	김용운 김용국교수의 수학클리닉	김용운, 김용국	김영사	10
	우리 문화의 수수께끼 1-2	주강현	한겨레신문사	23
총합	13종			177레코드



을 계산하였다. 항목별 입력률, 즉 필드별 입력 현황은 <표 3>과 같다.

서명, 저자, 출판사, 분류번호, 가격, 발행일 등의 입력률은 높으나 그 외 항목은 입력률이 매우 낮았다. 본서명과 저자사항은 모든 입력자가 반드시 입력하고 있으며 출판사는 99%, 분류번호는 86% 수준으로 입력되었다. 반면 ISBN이나 표지이미지, 페이지수, 판사항은 중복레코드 가운데서 적합한 레코드를 식별할 수 있게 도와주는 중요한 요소임에도 불구하고 입력률이 각각 40%, 29%, 46%, 0%에 불과하였다. 이용대상과 간략내용은 독서 관련 정보를 체계적으로 활용해야할 학교도서관 데이터베이스의 특성상 어렵더라도 책임감을 가지고 입력해야할 항목인데 36%와 22%에 불과하였다.

완전성은 레코드에 포함된 데이터의 내용이

원자료의 중요한 속성을 상세히 표현하고 있는 정도로서 각 서지에 대해 기술해야할 항목의 총수 가운데 실제로 기술된 항목수의 비율을 측정하였고 <표 4>에 나타내었다. 177개 레코드의 기술을 위해 실제로 사용된 항목수를 조사한 결과 총 1,867항목으로 하나의 레코드 당 평균 10.55항목이 입력되었음을 알 수 있었다. 그러나 이상적인 레코드 구성을 위해 입력해야만 할 항목수를 조사한 결과는 총 3,879항목으로 하나의 레코드 당 평균 21.92항목이었다. 따라서 조사대상 레코드의 완전성은 48.12%에 불과하였다.

하나의 서지를 기술하기 위해 사용할 수 있는 데이터 항목은 최대 34개이다. 그러나 현실적으로 서지기술을 위해 34개 항목을 모두 사용할 필요는 없기 때문에 본 연구에서는 34개

<표 3> 서지레코드의 필드별 입력률

항 목	입력된 레코드수	입력률	항 목	입력된 레코드수	입력률	항 목	입력된 레코드수	입력률
본서명	177	100%	권차서명	8	47%	기타서명	18	17%
저자	177	100%	페이지수	82	46%	목차	6	3%
출판사	176	99%	추가저자	33	42%	주기	8	? *
분류번호	153	86%	추가저자역할어	33	42%	언어	0	0%
가격	150	85%	ISBN	71	40%	생물년	0	0%
발행일	138	78%	원서명	14	37%	검색에서 이용될 키워드	0	0%
권차	25	63%	이용대상	63	36%	판사항	0	0%
발행지	106	60%	유포된 상태	62	35%	저작권	0	0%
크기	96	54%	삽도	33	29%	연관저록	0	0%
저자역할어	89	50%	표지이미지	51	29%	URL	0	0%
총서사항	24	49%	ISBN(세트)	1	25%			
분류명	85	48%	간략내용(초록)	39	22%			

\* 주기사항은 자료의 내용에 대해 설명이나 보완을 요하는 경우 기술하는 요소로 입력여부는 입력자의 재량에 따라 변동이 크기 때문에 이 항목을 필요로 하는 레코드 총 수를 명확히 할 수 없어 입력률을 계산하지 않았음.

〈표 4〉 레코드의 완전성

		합계	평균항목수	완전성 ( $A' + B' + C' / A + B + C$ $A' / A, B' / B, C' / C$ )
기술된 항목수	항목총수( $A' + B' + C'$ )	1,867	$1867/177 = 10.55$	$10.55/21.92 \times 100 = 48.12\%$
	기본사항수( $A'$ )	1,444	$1444/177 = 8.16$	$8.16/13.32 \times 100 = 61.26\%$
	부가적 기본사항수( $B'$ )	156	$156/177 = 0.88$	$0.88/2.53 \times 100 = 34.78\%$
	확장사항수( $C'$ )	267	$267/177 = 1.51$	$1.51/6.07 \times 100 = 24.87\%$
기술할 항목수	항목총수( $A + B + C$ )	3,879	$3879/177 = 21.92$	
	기본사항수( $A$ )	2,358	$2358/177 = 13.32$	
	부가적 기본사항수( $B$ )	447	$447/177 = 2.53$	
	확장사항수( $C$ )	1,074	$1074/177 = 6.07$	

항목을 기본사항, 부가적 기본사항, 확장사항 등 3그룹으로 나누어 다시 집계해 보았다.

기본사항(A)에 포함될 항목은 단행본의 기본사항 입력화면에서 제시한 ISBN, 분서명, 권차, 권차서명, 저자, 역할어, 출판사, 발행일, 페이지수, 크기, 판사항, 가격, 본문기술언어, 이용대상, 분류번호, 간략내용(초록) 등 16항목으로 제한하였고, 소장/확장서지정보 입력 화면에서 제시한 항목 중 ISBN(세트), 총서명, 부서명, 원서명, 추가저자, 추가 저지역할어, 삽도 등 해당 저작에서 필수적으로 추가 기술되어야 할 7항목은 부가적 기본사항(B)으로 조사하였다. 이 외에 생몰년, 발행지, 유포된 상태, 저작권, 분류명, 검색에서 이용될 키워드, 주기, 연관저록, URL, 목차, 표지이미지 등 11항목은 확장사항(C)으로 조사하였다.

집계 결과 조사대상 177개 레코드에 대해 이상적인 레코드 구성을 위해 입력해야만 할 기본항목의 수는 레코드 당 평균 13.32개, 부가적 기본항목의 수는 레코드 당 평균 2.53개, 확장항목의 수는 레코드 당 평균 6.07개였다.

그러나 실제로 기술된 항목수는 기본항목이 레코드 당 평균 8.16개, 부가적 기본항목은 레코드 당 평균 0.88개, 확장항목은 레코드 당 평균 1.51개였다. 따라서 기본항목의 완전성 정도는 입력해야 할 기본항목 13.32개 중 실제로 8.16개가 기술되었으므로 약 61.26%로 측정되었고, 부가적 기본항목의 완전성 정도는 2.53개 중 0.88개가 기술되어 약 34.78%, 확장항목의 완전성 정도는 6.07개 중 1.51개가 기술되었으므로 약 24.87%로 나타났다. 세 그룹을 통합한 완전성 정도는 약 48.12%이다.

#### 4. 4 정확성

정확성 측정에 사용된 자료는 완전성 측정에 사용된 자료와 동일한 177개 레코드의 상세 목록이다. 서지기술에 있어 필수항목이라 할 수 있는 기본사항의 입력률이 61.26%이기는 하나 기술된 항목의 내용을 분석해보면 상당히 심각한 문제점을 발견하게 된다. 오류 없이 입력된 레코드 수는 177개 레코드 중 14개에 불

과하여 7.9%인 것으로 드러났고 92%에 해당하는 163건에서 크고 작은 오류들이 발견되었다. 이런 오류들로 인해 동일한 저작에 대해 기술한 레코드들이 하나로 통합되지 못한 채 각각 별개의 레코드로 저장되어 결국 상이한 저작을 표현한 레코드로 처리되고 마는 것이다.

기술된 데이터가 실제 값과 동일한지 여부를 파악하기 위해 각 항목별로 기술 내용의 철자나 데이터값 오류, 누락 여부, 혹은 입력 규칙 준수 여부를 조사하였다. 기본사항과 부가적 기본사항에 속하는 중요 필드를 중심으로 오류발생 사례를 집계해 발생을 순으로 <표 5>에 정렬하였다.

가장 빈번하게 발생하는 오류는 저자가 2인 이상일 때 공저 역할어를 누락하는 것으로 발생율이 100%였다. 발생율이 50%~78.8%에 해당하는 내용들은 기타서명(부서명, 잡제, 관칭, 관제) 누락, 삽도표시 누락, 원서명 누락, ISBN(개별번호) 누락, 제2저작자 혹은 제3저작자(역자, 그림) 누락, 페이지수 누락, 본서명에서 한자로 표기할 내용을 한글로 표기, 권차서명 누락, 총서명 누락, 총서번호 누락, 저자(1인) 역할어 누락 등이다. 이 외에 50%에는 못 미치지만 중요한 사항으로 발행년을 판단하는 데 있어 판의 변화 없이 인쇄년만 달라졌는데도 발행일 항목에 재쇄년을 입력하는 오류가 빈번히 발생하여 발생을 46.3%에 달하였다. 또 판사항 누락, 서양저자명 입력 시 이름과 성 사이를 띄어 쓰도록 규정하였는데 띄어 쓰지 않은 오류, 원서명 입력 시 대문자표기규칙을 지키지 않은 오류 등이 발견되었다.

오류의 내용을 보면 입력규칙에 맞지 않아

발생한 오류보다는 누락 오류가 주류를 이루었다. 기술된 서지 내용이 워낙 간단하기 때문에 규칙 위반 오류의 건수가 상대적으로 적게 나오는 것으로 보인다.

이렇듯 빈번히 발생하는 오류의 원인을 좀 더 자세히 살펴보고 오류를 방지하기 위해서는 어떤 점에 유의해야 하는지 파악하기 위해 각 레코드에서 발생하는 오류들을 주요 입력요소 별로 분류하고 각 필드의 입력규칙과 비교하여 분석해보았다.

#### (1) ISBN 입력의 오류

국제표준도서번호인 ISBN은 도서자료를 개별화하기 위해 출판 전에 해당 출판물에 부여하는 10자리 숫자의 고유번호로, 도서마다 부여되는 일련번호 10자리를 ‘-’ 없이 입력한다. ISBN은 서지를 구분할 수 있는 책의 주민등록번호와 같은 역할을 하므로 목록 기입시 반드시 기재하도록 DLS 목록입력규칙(단행본)에서 규정하고 있다. 그러나 <표 6>에서 보듯이 본 연구를 위해 선정된 레코드 중 ISBN을 기재하지 않은 비율은 59.9%에 달하여 각 서지를 식별하는 데 있어 중요한 역할 자료로서의 ISBN에 대한 인식율은 매우 낮음을 알 수 있다. ISBN은 종합목록으로부터 레코드 반입시 유사한 레코드들로부터 적합한 레코드를 판별해 내는 데에 매우 유용한 요소이고 또 입력시 복잡한 규칙을 따라야 하는 요소도 아니므로 반드시 입력해야 할 필수요소라는 인식을 가져야겠다.

#### (2) 서명사항 입력의 오류

서명사항에 대한 인식은 매우 단순하여 본

〈표 5〉 주요 필드의 오류 발생률

오류의 내용	오류레코드수	해당도서수	발생율	해당사항
저자(2인 이상) 공저 역할어 - 누락	10	10	100.0%	저자사항
기타서명(부서명, 잡제, 관칭, 관제) - 누락	86	109	78.8%	서명사항
삽도표시 - 누락	81	114	71.0%	형태사항
원서명 - 누락	24	38	63.1%	서명사항
ISBN - 각 책의 개별번호 누락	106	177	59.9%	ISBN
제2, 3저작자(역자, 그림) - 누락	37	68	54.4%	저자사항
페이지수 - 누락	94	177	53.1%	형태사항
본서명 - 한자를 한글로 표기	9	17	52.9%	서명사항
권차서명 - 누락	9	17	52.9%	서명사항
총서명 - 누락	25	49	51.0%	총서사항
총서번호 - 누락	25	49	51.0%	총서사항
저자(1인) 역할어 - 누락	84	167	50.2%	저자사항
발행년 - 판단 오류(판에 변화없는데 재쇄년을 입력, 초판년이전 잘못입력)	82	177	46.3%	발행사항
크기 - 누락	80	177	45.1%	형태사항
판사항 - 누락	9	23	39.1%	판사항
발행지 - 누락	69	177	38.9%	발행사항
본서명 - 띄어쓰기 오류	61	177	34.4%	서명사항
서양저자명 - 이름 성 사이 띄어쓰기 오류(띄어쓰지 않음)	11	38	28.9%	저자사항
권차표시 - 표지와 다름	9	40	22.5%	서명사항
원서명 - 대문자표기규칙 어긋(첫 단어의 두문자만 대문자 입력인데 모두 대문자로 입력, 고유명사의 두문자를 대문자로 표기안함, 불규칙)	8	38	21.0%	서명사항
발행년 - 누락	30	177	16.9%	발행사항
가격정보 - 누락	24	177	13.5%	가격정보
권차표시 - 누락	4	40	10.0%	서명사항
동양저자명 - 이름 성 사이 띄어쓰기 오류(띄어쓰) (일서30건중 발생율은 33.3%)	10	139	7.1%	저자사항
본서명 - 철자 오류	10	177	5.6%	서명사항
본서명 - 일부 입력, 추가 입력	9	177	5.0%	서명사항
저자명 - 철자 오류	6	177	3.3%	저자사항
역할어 앞 띄어쓰기 - 오류	6	177	3.3%	저자사항
저자명 - 일부 입력, 추가 입력	4	177	2.2%	저자사항
크기 - 오류	4	177	2.2%	형태사항
총서명 - 철자 오류	1	49	2.0%	총서사항
발행처 - 판단 오류	3	177	1.6%	발행사항
번역자 - 선택 오류	1	68	1.4%	저자사항
SBN - 입력 오류	2	177	1.1%	ISBN
페이지수 - 계산 오류	2	177	1.1%	형태사항
본서명 - 기타 서명과 혼동	1	177	0.5%	서명사항
발행지 - 입력 오류	1	177	0.5%	발행사항
발행처 - 누락	1	177	0.5%	발행사항

〈표 6〉 ISBN 입력시 오류 내용

오류의 내용	오류레코드수	해당도서수	발생율	해당사항
ISBN - 각 책의 개별번호 누락	106	177	59.9%	ISBN
ISBN - 입력 오류	2	177	1.1%	ISBN

서명은 100% 입력되었으나 본서명 이외에 권차, 권차서명, 원서명, 부서명, 잡제, 총서명 등을 입력하려는 노력은 부족한 것으로 보인다.

본서명도 입력률은 100%이나 정확성을 판단해볼 때는 다소 문제점이 드러났다. 본서명은 출판물의 주된 서명으로 권차를 뺀 순수한 도서명만 입력하며 실물자료의 표제지, 표지, 관련지순으로 참조하여 입력해야 한다. 그러나 입력된 본서명을 검토한 결과 실물자료에 나타난 그대로 입력되지 않고 탈자가 있거나 ('나, 이제 외톨이와 안녕할지 몰라요'를 '나 이제 외톨이와'로 입력) 불필요한 문자가 추가 입력된 경우('얼굴 빨개지는 아이'를 '얼굴이 빨개지는 아이'로 입력), 철자 오류('그 많던 싱아는 누가 다 먹었을까'를 '그 많던 상아는 누가 다 먹었을까'로 입력), 한자가 한글로 표기된 경우('대한민국史'를 '대한민국사'로 입력), 띄어쓰기 오류('나, 이제 외톨이와 안녕할지 몰라요'를 '나이제외톨이와안녕할지몰라요'로 입력) 등 사소한 오류 외에도 총서명이나, 부서명 등 기타서명과 혼동하여 입력된 경우('7년간의 실수'를 총서명인 '국어교사모임의글쓰기교육'으로 입력)도 발견되었다.

권차서명은 다권본의 경우 본서명 외에 권차에 딸린 서명으로 권차서명란에 따로 기재하도록 규정되어 있다. 권차는 두 개 이상의 권수로 형성된 단위물에서 그 순서를 표시한

숫자나 단어로 권차란에 기재해야 한다. 그러나 권차표시를 누락하거나 실물표지와 다른 방식으로 입력한 경우('01'을 '1'로 입력)가 다소 있었고 권차표시를 입력했다 하더라도 권차서명을 누락한 경우는 더 많았다. 따라서 본서명으로 검색하여 다권본이 출력되었을 때 권차표시 없이 출력되는 경우가 많아 이용자가 원하는 자료인지를 정확히 파악하기 어려운 것으로 드러났다.

원서명은 번역서의 경우 원래 저작의 원어로 된 서명으로 확장서지정보에 입력해야하나 누락된 경우가 63.1%에 달하여 한글로 된 서명 이외의 것에는 큰 관심을 두지 않는 것으로 보인다. 또한 원서명을 입력할 때 대문자 처리에 일관성이 결여되어 있음을 발견하였다. AACR2에 따르면 표제의 첫글자를 대문자로 표기하고, 개인명이나 단체명 등의 고유명사는 관사, 전치사 접속사를 제외한 각각의 분리된 단어나 두문자마다 대문자로 표기할 것을 규정하고 있다. 그러나 원서명을 모두 대문자로 입력하거나(L' ARBRE DES POSSIBLES) 불규칙하게 대소문자를 섞어 쓰는 경우(J'aime pas Mourir)가 있었고 고유명사의 두문자를 대문자로 표기하지 않은 경우(Marcellin caillou)도 있었다.

그 외에 본서명을 설명하거나 보완하는 부차적 서명인 부서명을 포함하여 잡제, 관칭, 관제 등을 누락한 경우를 모두 집계하면 무려

78.8%에 달하여 본서명 이외의 정보에 대해서는 입력의 필요성을 크게 느끼지 않고 있는 것으로 보인다.

서명사항과 관련된 오류들은 조금만 주의를 기울이면 쉽게 개선할 수 있는 내용이 대부분 이었고 원서명 입력을 위해 대문자표기규칙을 좀 더 주의해서 숙지해야 할 필요성이 있었다. 서명사항 입력 시 발생하는 오류는 <표 7>에 정리하였다.

### (3) 저자사항

저작의 지적인 또는 예술적인 창작에 있어 책임이 있거나 기여한 개인이나 단체를 표시하는 저자는 표지 또는 판권지에 기입된 그대로 입력한다. 동양저자(한국인, 중국인, 일본인 등 성, 이름 순으로 되어있는 경우)에 대해서는 띄어쓰기 없이 기재하고, 서양저자의 경우 저자명을 도치하지 말고 이름과 성 사이에 띄어쓰기를 하여 기재하도록 규정하고 있다. 저, 편, 역 등 저자역할어는 선택항목에서 선

택하며 선택항목에 없는 역할어는 직접 기재하도록 하고 있다.

저자명을 검토한 결과 탈자가 있거나(‘하이타니겐지로’를 ‘하이타니겐지’로 입력) 불필요한 문자가 추가 입력된 경우(‘주강현’을 ‘주강현지음’으로 입력, ‘지음’은 저자 항목이 아닌 역할어 항목에서 선택하거나 직접 기재해야 함), 철자 오류(‘하이타니겐지로’를 ‘하아타니겐지로’로 입력), 띄어쓰기 오류 등이 발견되었다. 서양저자명 입력 시 띄어쓰기를 하지 않은 오류(‘조지 오웰’을 ‘조지오웰’로 입력)가 많았고 동양저자명 입력 시 띄어쓰기를 한 오류(‘구로야나기테츠코’를 ‘구로야나기 테츠코’로 입력)도 발견되었다. 특히 일본저자의 경우 성과 이름을 띄어 쓴 오류가 많았다. 저자명 입력시 띄어쓰기에 주의해야 하는 이유는 저자기호가 성을 기준으로 작성되기 때문이다. DLS 활용 매뉴얼 뿐 아니라 목록구축가이드에도 ‘저작자를 입력할 때 서양인 성명은 표제지나 판권지에 찍여 있는 그대로 입력하면 되지만,

<표 7> 서명사항 입력 시 오류 내용

오류의 내용	오류레코드수	해당도서수	발생율	해당사항
본서명 - 일부 입력, 추가 입력	9	177	5.0%	서명사항
본서명 - 철자 오류	10	177	5.6%	서명사항
본서명 - 한자를 한글로 표기	9	17	52.9%	서명사항
본서명 - 기타 서명과 혼동	1	177	0.5%	서명사항
본서명 - 띄어쓰기 오류	61	177	34.4%	서명사항
권차표시 - 누락	4	40	10.0%	서명사항
권차표시 - 표지와 다름	9	40	22.5%	서명사항
권차서명 - 누락	9	17	52.9%	서명사항
원서명 - 누락	24	38	63.1%	서명사항
원서명 - 대문자표기규칙 어긋	8	38	21.0%	서명사항
기타서명(부서명, 잡제, 관청, 관제) - 누락	86	109	78.8%	서명사항

저자기호 자동 생성 시 오류를 막기 위해 한국인, 일본인과 중국인 성명은 띄우지 않고 모두 붙여서 표기해야 한다'고 띄어쓰기 지침이 세밀하게 제시되어 있다(한국교육학술정보원 2003a; 한국교육학술정보원 2003b).

번역과 그림에 기여한 제2, 제3저작자를 누락한 경우는 무려 54.4%에 달해 저자명은 하나만 입력해도 충분하다는 인식을 하고 있는 것으로 보인다. 저자역할어 입력률도 매우 낮은 것으로 드러났다.

저자사항 입력 시에도 사소한 실수를 하지 않도록 좀더 주의를 기울여야 하겠고 특히 띄어쓰기 문제는 동양서와 서양서의 경우를 확실히 구분해서 규칙을 익힐 필요가 있겠다. 저자사항 입력 시 발생하는 오류는 <표 8>에 정리하였다.

(4) 발행사항과 판사항

발행지는 출판자의 회사가 위치한 도시명이며, 출판사는 출판과 배포, 공개발행의 책임을 진 개인이나 단체로서 출판사 앞에 붙는 법인

형태, 업종명(예: 주식회사, (주)도서출판, 사단법인)은 생략한다. 발행년은 해당판의 인쇄년도가 아닌 출판년도를 입력한다. 즉 같은 도서의 연속된 인쇄(판개념이 아닌 쇄개념)일 경우 초판년도를 기재하고 인쇄년도는 소장정보에서 기재해야한다.

발행지를 누락하거나 잘못 입력하는 경우('서울'을 '파주'로 입력)가 많았고 출판사를 누락하거나 잘못 입력하는 경우('웅진닷컴'을 '웅진닷컴'으로 입력)도 있었다. 발행년을 누락하는 경우도 다소 있었지만 무엇보다 눈에 띄는 것은 발행년의 판단 오류이다. 판에 변화가 없이 인쇄년도만 다른데도 발행년 항목에 재쇄년을 입력하는 경우가 46.3%나 발생하여 정확한 입력지침을 교육시킬 필요성이 있었다.

판사항이란 도서의 판차나 판의 종류를 기술하는 것으로 개정판, 증보판, 한정판, 축소판, 아시아판 등의 표시이며 쇄차는 판표시에서 제외하고 초판에 대한 표시는 하지 않는다. 본 연구에서는 '컬러개정판'임을 명기해야 할 경우가 9건이나 있었는데도 모두 누락하여

<표 8> 저자사항 입력 시 오류 내용

오류의 내용	오류레코드수	해당도서수	발생율	해당사항
동양저자명 - 이름 성 사이 띄어쓰기 오류 (띄어쓰)(일서 30종 중 10건 오류)	10	139	7.1%	저자사항
서양저자명 - 이름 성 사이 띄어쓰기 오류 (띄어쓰지 않음)	11	38	28.9%	저자사항
저자명 - 일부 입력, 추가 입력	4	177	2.2%	저자사항
저자명 - 철자 오류	6	177	3.3%	저자사항
번역자 - 선택 오류	1	68	1.4%	저자사항
제2, 3저작자(역자, 그림) - 누락	37	68	54.4%	저자사항
저자(1인) 역할어 - 누락	84	16.7	50.2%	저자사항
저자(2인 이상) 공저 역할어 - 누락	10	10	100%	저자사항
역할어 앞 띄어쓰기 - 오류	6	177	3.3%	저자사항

39.1%로 집계되었다.

발행사항과 관련하여 유의할 점은 발행년의 판단과 관련하여 정확한 지식이 필요하다는 점이다. 이는 어려운 판단이 요구되는 지식도 아니고 판의 변화가 없을 때에는 반드시 초판년도를 기재하고 인쇄년도는 소장정보에 기재해야한다는 점만 기억하면 되는 것이다. 판사항 역시 무심코 건너 띄어도 무관할 만큼 사소한 정보가 아니며 특히 학술자료인 경우 중요한 정보를 제공하는 경우가 많다는 사실을 알아야 할 것이다.

발행사항과 판사항 입력 시 발생하는 오류는 <표 9>에 정리하였다.

(5) 형태사항  
페이지수에는 인쇄된 최종면수를 기재하고

면수표시가 없는 책은 직접 면수를 세어 각괄호 [ ]p. 안에 표기한다. 크기는 자료의 표지나 외형의 세로길이를 cm단위로 올림하여 기재한다. 삽도는 본문 내용을 보완하고, 명료하게 하기 위해 출판물에 삽입된 그림이나 도표, 지도, 그래프, 사진 등을 포함한다.

페이지수 뿐 아니라 크기, 삽도표시 등 형태사항에 대해서는 전반적으로 누락율이 높은 편이었고 페이지수 계산 오류도 있었다.

페이지수는 중복레코드가 많이 검색되었을 때에 적합 레코드 판단에 유용하게 쓰이는 항목이고 또 어려운 판단이 요구되는 항목이 아니다. 따라서 필히 입력해야할 기본사항이라는 점이 좀 더 강조되어야 할 것이다.

형태사항 입력 시 발생하는 오류는 <표 10>에 정리하였다.

<표 9> 저자사항 입력 시 오류 내용

오류의 내용	오류레코드수	해당도서수	발생율	해당사항
발행지 - 누락	69	177	38.9%	발행사항
발행지 - 입력 오류	1	177	0.5%	발행사항
발행처 - 누락	1	177	0.5%	발행사항
발행처 - 입력 오류	3	177	1.6%	발행사항
발행년 - 누락	30	177	16.9%	발행사항
발행년 - 판단 오류 (판에 변화없는데 재쇄년을 입력)	82	177	46.3%	발행사항
판사항 - 누락	9	23	39.1%	발행사항

<표 10> 형태사항 입력 시 오류 내용

오류의 내용	오류레코드수	해당도서수	발생율	해당사항
페이지수 - 누락	94	177	53.1%	형태사항
페이지수 - 계산 오류	2	177	1.1%	형태사항
크기 - 누락	80	177	45.1%	형태사항
크기 - 오류	4	177	2.2%	형태사항
삽도표시 - 누락	81	114	71.0%	형태사항



(6) 총서사항 및 가격정보

여러 개로 분리된 출판물이 상호 연결되어 있을 때 각 출판물이 자체의 본서명을 지님과 동시에 그 무리 전체에 통용되는 집합적인 서명을 갖고 있으면 그것을 총서명에 기재하게 된다. 본 조사에서 총서명과 총서번호가 누락되는 정도는 50%선으로 서명사항에서 보았듯이 본서명 이외의 정보에 대해서는 입력의 필요성을 크게 느끼지 않고 있는 것으로 보인다.

반면 가격정보의 누락율은 13.5%에 불과하여 입력자들이 가격정보를 매우 중요한 사항으로 인식하고 있음을 살펴볼 수 있다. 다만 가격정보는 도서의 정가와 유포된 상태를 입력해야 하는데, 할인가나 구입가를 입력하고 유포상태에 대해서는 누락하는 경우가 종종 있었다.

총서사항 및 가격정보 입력 시 발생하는 오류는 <표 11>에 정리하였다.

5. 결론

각급 학교도서관들이 공동으로 목록을 구축하고 활용함으로써 목록작업을 획기적으로 개선할 수 있다는 원래의 취지와 달리 DLS는 양적인 성장을 뒷받침하지 못하는 조악한 데

이터의 품질과 시스템의 불안정으로 인해 난항을 겪고 있다. 목록작업이 용이해지기를 기대했던 도서관 실무자들은 공동목록 검색 시 출력되는 수많은 중복레코드들의 내용을 일일이 확인하려면 더 많은 시간이 낭비되기 때문에 차라리 신규입력을 해버리는 경우가 많아 결과적으로 종합목록 품질은 더 나빠질 가능성이 높고 결국 DLS 사용 저조로 이어지는 현상이 나타나고 있다.

본 연구는 종합목록데이터베이스의 품질관리의 시급성을 인식하여 행하여진 일차 작업으로서 부산광역시 교육청 DLS 레코드의 품질을 분석하였고 그 결과를 종합하여 품질관리를 위한 제언을 하고자한다. 특히 각 필드에서 발견되는 오류의 원인을 상세히 살펴보고 이를 기초로 오류 방지를 위해 입력 실무자들이 유의할 사항을 제시하고자 한다.

본 연구에서 분석한 내용을 요약하면 다음과 같다.

첫째, DLS 데이터베이스의 적중률은 98%~100%로 매우 높았고 중학생을 위한 추천 자료의 범위를 충분히 만족시키는 수준인 것으로 나타났다.

둘째, 동일 저작에 대해 상이한 레코드가 다수 존재하는 중복성은 7.46으로 심각한 수준이었으며 이들이 중복레코드임에도 불구하고

<표 11> 총서사항 및 가격사항 입력 시 오류 내용

오류의 내용	오류레코드수	해당도서수	발생율	해당사항
총서명 - 누락	25	49	51.0%	총서사항
총서번호 - 누락	25	49	51.0%	총서사항
총서명 - 철자 오류	1	49	2.0%	총서사항
가격정보 - 누락	24	177	13.5%	가격정보

고 별개의 레코드로 처리된 이유는 각 학교도서관에서 입력한 레코드의 구조에 일관성이 없고 입력된 내용이 정확하지 않기 때문이다.

셋째, 레코드의 완전성 정도는 기본항목 61.26%, 부가적 기본항목 34.78%, 확장항목 24.87%이며, 통합할 경우 48.12%에 그쳐 입력해야 할 항목에 대한 인식이 매우 상식적인 수준에 그치고 있음을 알 수 있다.

넷째, 레코드의 정확성 분석 결과 92%에 해당하는 레코드에서 크고 작은 오류가 발견되었다. 이런 오류들로 인해 동일한 저작에 대한 레코드들이 하나로 통합되지 못하고 별개의 레코드로 저장되어 중복성을 높이는 요인이 되고 있었다. 해당 항목을 입력하지 않고 무책임하게 누락해버린 오류가 높은 비중을 차지하고 있었고 입력규칙을 준수하지 않거나 데이터 값을 잘못 입력한 비중은 상대적으로 낮았다. 입력규칙 위반 오류들도 어려운 판단을 요구하는 규칙이 아니라 부주의와 안이한 태도 때문에 발생한 오류들이 많았다. 본서명 입력 시 한자를 한글로 표기한 오류, 발행년 입력 시 판의 변화가 없는데 재쇄년을 입력한 오류, 본서명과 저자명 입력 시 띄어쓰기 오류, 원서명 입력 시 대문자 표기 오류, 철자 오류, 탈자 혹은 추가 입력 오류 등이 주로 발견되었다.

이와 같은 분석 결과를 종합하여 데이터베이스의 품질관리를 위한 제언을 하고 특히 오류 방지를 위해 입력 실무자들이 유의할 사항을 제시하고자 한다.

첫째, 데이터의 심각한 중복성과 낮은 정확률은 DLS 데이터베이스 구축 초기부터 품질관리 대책을 철저히 세우지 않았던 데에 기인

한다. 대부분의 학교도서관은 담당사서가 없는 상태에서 아르바이트를 고용하거나 외부 용역에 의존하여 양 위주의 입력 작업을 실시하였고 입력데이터에 대한 품질 검증 작업에 소홀할 수밖에 없는 실정이다. 양질의 데이터를 위해 일차적으로는 각급 학교도서관에서 보다 철저한 검증작업을 거쳐야 하겠지만 열악한 학교도서관의 현실을 감안할 때 획기적인 지원책이 없는 한 큰 기대를 하기 어려운 것으로 보이며 오히려 센터 측에서 관리 기능을 강화하는 것이 현실적인 대책이라고 할 것이다. 또 현재 시스템에 중복체크 기능이 있다고는 하나 그 기능이 만족할만한 수준이 아니고, 센터관리자가 중복 데이터의 서지병합 작업을 통해 종합목록의 품질관리를 하는 것 역시 수작업으로 해결하기에는 오류데이터의 양이 너무나 많다. 따라서 오류 및 중복체크 기능을 좀 더 세밀하게 재검토하여 전문적인 품질관리프로그램을 가동하는 자동화 방안을 제안한다.

둘째, 완전성이 지극히 낮은 이유는 지금까지 개별 학교에서 구축하여 업로드한 목록이 저자, 서명, 출판사 등의 단편적 정보만 입력된 수준 이하의 레코드이기 때문이다. 따라서 필수적으로 입력해야 할 항목들을 지정하고 최소 입력 수준을 지키지 않으면 공동목록 데이터베이스로 업로드 되지 않도록 하는 강한 규제가 있어야 할 것이다. 특히 ISBN, 표지이미지, 페이지수 항목은 중복 레코드 가운데서 적합 레코드를 판단할 수 있는 단서가 되기 때문에 가능한 입력해야 하겠고 이용대상과 간략내용 항목은 도서관 활용수업과 독서교육이라는 전문성을 확보해야 할 학교도서관용 전

문 데이터베이스임을 고려할 때 책임감을 가지고 입력해야 할 사항이라 하겠다.

셋째, 학교도서관의 입력 담당자는 최소한 DLS 활용 매뉴얼의 부록에 제시된 5쪽짜리 간략판 목록입력규칙만이라도 눈여겨보기를 제안한다. 본 연구에서 발견한 오류에 대한 해법은 이 5쪽짜리 간략판 안에 대부분 포함되어 있었다. 입력자가 문헌정보학 전공자가 아니고 전문지식이 부족하다는 이유보다는 목록 구축 작업이 가볍게 해도 되는 일이라는 인식이 현재의 학교도서관 상황에서 더 치명적인 것이며 이러한 인식이 존재하는 한 조악한 레코드는 꾸준히 생성될 수밖에 없을 것이다. 입력자들은 최소 입력 수준을 지킨다는 기본 원칙 하에 작업을 하되 사소한 실수를 하지 않도록 좀더 주의를 기울여야 하겠고 아래에 제시하는 유의사항만이라도 숙지해야 하겠다.

- 1) ISBN은 적합한 레코드를 판별해 내는데 매우 유용한 요소이므로 반드시 입력해야 할 필수요소라는 인식을 가져야겠다.
- 2) 원서명 입력을 위해 대문자표기규칙을 좀 더 주의해서 숙지해야 할 필요성이 있다.
- 3) 저자사항 입력 시 띄어쓰기 문제는 동양서와 서양서의 경우를 확실히 구분해서 규칙을 익힐 필요가 있다.
- 4) 발행년 입력 시 판의 변화가 없을 때에

는 반드시 초판년도를 기재하고 인쇄년도는 소장정보에 기재해야한다는 점을 기억해야 하겠다.

- 5) 판사항은 사소한 정보가 아니며 특히 학술자료인 경우 중요한 정보를 제공하는 경우가 많으므로 해당시 반드시 입력해야 할 것이다.
- 6) 페이지수는 중복레코드가 많이 검색되었을 때에 적합 레코드 판단에 유용하게 쓰이는 항목이므로 필히 입력해야할 기본사항이라는 점이 좀 더 강조되어야 할 것이다.

본 연구의 조사 대상 데이터가 많지 않아 세밀히 살펴보지 못한 항목들이 아직도 많이 있지만 최소입력수준에서 기본 항목에 대해 요구되는 주의사항은 대략적으로 살펴보았다. 일선 담당자들은 본 연구에서 지적한 내용을 우선적으로 유념하고 나머지 항목에 대해서도 입력규칙을 일람하는 정도의 수고를 아끼지 말아야 할 것이며, 종합목록을 관리하는 센터 측에서도 입력 데이터에 대한 평가와 관리가 부족했었다는 점을 인정하고 추후 목록입력규칙을 준수하도록 만드는 장치, 품질관리를 위한 전문 인력 보강, 양질의 외부 목록 데이터 확보, 데이터베이스의 수준 유지를 위한 제도적 장치를 조속한 시일 내에 실현해야만 할 것이다.

## 참 고 문 헌

- 교육인적자원부, 한국교육학술정보원. 2003. 『2003 디지털자료실지원센터 구축·운영 시·도 관계자 협의회 자료』. [cited 2004.4.10].  
<<http://dls.busanedu.net>>.
- 김상준. 2003. 의학 분야 Web DB의 품질 평가: PubMed와 Embase를 대상으로. 『한국문헌정보학회 학술발표논문집』, 17: 33-56.
- 김석영. 2002. 과학기술분야 해외 온라인 데이터베이스 평가 연구. 『정보관리연구』, 33(1): 61-80.
- 김성희. 2003. 데이터베이스 평가모델 개발을 위한 해외사례 조사연구. 『국회도서관보』, 40(2): 15-33.
- 노지현. 2003. 유용성의 관점에서 본 도서관목록의 품질. 『한국문헌정보학회지』, 37(2): 107-134.
- 윤정우. 2003. 연속간행물 종합목록 데이터베이스의 레코드 품질 평가. 『한국문헌정보학회지』, 37(1): 27-42.
- 이만수, 이상열, 신인수. 2003. 서지 데이터베이스의 품질평가에 관한 연구. 『정보관리연구』, 34(3): 21-39.
- 이병기. 2003. 인터넷 기반 학교도서관 정보시스템의 평가에 관한 연구. 『정보관리학회지』, 20(1): 31-51.
- 이응봉, 조현양, 류범중, 최재황. 2001. 과학기술분야 데이터베이스의 품질향상을 위한 품질평가 연구. 『한국문헌정보학회지』, 35(2): 109-132.
- 이제환. 1998. 분산체제로 구축된 통합 DB의 품질관리에 관한 연구. 『한국문헌정보학회지』, 32(3): 179-206.
- \_\_\_\_\_. 2002. 공동목록 DB의 품질평가와 품질관리. 『한국문헌정보학회지』, 36(1): 61-89.
- 조순영. 2002. 종합목록 데이터의 오류 유형에 관한 연구 -KERIS 종합목록의 학위논문 서지데이터를 중심으로-. 『한국문헌정보학회지』, 36(4): 5-19.
- 학교도서관을 살리는 교사들. [online]. [cited 2004.8.5].  
<<http://cafe.daum.net/libte>>.
- 한국과학기술정보연구원. 2001. 『과학기술 전문정보 DB 통합 및 표준화 방안 연구』. 서울: 한국과학기술정보연구원.
- 한국교육학술정보원. 2002. 『해외학술DB 선정 및 평가모델 개발에 관한 연구』. 서울: 한국교육학술정보원. KR 2002-4.
- \_\_\_\_\_. 2003a. 『DLS 운영을 위한 목록 구축 가이드 -DLS I-』. 서울: 한국교육학술정보원. GM 2003-12.
- \_\_\_\_\_. 2003b. 『DLS 활용 매뉴얼 - DLS I 단위학교용 -』. 서울: 한국교육학술정보원. GM 2003-17.
- \_\_\_\_\_. 2003c. 『DLS 활용 매뉴얼 - DLS I 지원센터용 -』. 서울: 한국교육학술정보원. GM 2003-18.
- 한국데이터베이스진흥센터. 1996a. 데이터베이스 품질 평가. 『데이터베이스 월드』, (04): 51-55.

- \_\_\_\_\_. 1996b. 데이터베이스 품질 평가. 『데이터베이스 월드』, (05): 69-80.
- \_\_\_\_\_. 2000. 『데이터베이스 품질 평가 항목』. 서울: 한국데이터베이스진흥센터.
- 함명식. 2002. 부산광역시교육청 디지털자료실지원센터의 구축 사례 연구. 『정보관리학회지』, 19(3): 31-48.
- Anglo-American Cataloguing Rules*. 1998. 2nd ed. 1998 revision. Ottawa: Canadian Library Association.
- Basch R. 1990. "Measuring the Quality of Data: Report of the Fourth Annual SCOUG Retreat." *Database Searcher*, 6(8): 34-45.
- Fox, C. 1994. "The Notion of Data and Its Quality Dimensions," *Information Processing & Management*, 30(1): 9-19.
- Jasco, Peter. 1995. Testing the Quality of CD-ROM Database. in Basch, Reva. *Electronic Information Delivery*. Brookfield, VT: Gower.

K C I